IN-65
185003
45P

NASA CONTRACTOR REPORT

NASA CR-192523

PARAMETRIC ANALYSIS OF ATMOSPHERIC PROCESSES

By M. Elshamy
Department of Mathematical Sciences
The University of Alabama in Huntsville
Huntsville, Alabama 35899

August 1993

(NASA-CR-192523) PARAMETRIC ANALYSIS OF ATMOSPHERIC PROCESSES Final Report (Alabama Univ.) 45 p

N94-10757

Unclas

G3/65 0185003

Final Report

Prepared for
NASA-Marshall Space Flight Center
Marshall Space Flight Center, Alabama 35812

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1993	**Contractor Report - Final Report**

4. TITLE AND SUBTITLE

Parametric Analysis of Atmospheric Processes

5. FUNDING NUMBERS

Delivery Order #139 of NAS8-36955

6. AUTHOR(S)

M. Elshamy

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Mathematical Sciences
The University of Alabama in Huntsville
Huntsville, AL 35899

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

George C. Marshall Space Flight Center
Marshall Space Flight Center, AL 35812

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NASA CR- 192523

11. SUPPLEMENTARY NOTES

Prepared for Earth Science & Applications Division, Space Science Laboratory, Science & Engineering Directorate.
COR: C. Kelly Hill

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Subject Category: 65

Unclassified--Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The different phases of the space shuttle mission operations and system analyses are influenced by several dynamic atmospheric parameters. These parameters are random variables which are not independent. Probabilistic models incorporating dependence structure of the Markovian type are analyzed. The theory established could be used to predict a GO-NOGO decision in the different phases of a mission.

14. SUBJECT TERMS	15. NUMBER OF PAGES	16. PRICE CODE
Analyses, Atmospheric Parameters, Dependence, Markovian	45	NTIS

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	**Unclassified**	**Unclassified**	**Unlimited**

ACKNOWLEDGMENTS

I would like to express my appreciation to William W. Vaughan, Director of Research Institute, for his support. My thanks also to O. E. Smith and S. I. Adelfang of New Technology, Inc., for their helpful discussion.

TABLE OF CONTENTS

I. INTRODUCTION

The different phases of the space shuttle mission operations and system analyses are influenced by several random perturbations due to the dynamics of atmospheric processes. From the mission planning point of view, there are few atmospheric conditions of interest, such as thunderstorm, precipitation, cloud ceiling, peak surface wind speed, etc. These atmospheric conditions, called parameters, are actually constraints on the mission operations. An atmospheric parameter is a random variable which attains either a permissible or not permissible value. As such, each of these atmospheric parameters is assigned the values of either 0 or 1, for GO or NOGO outcomes, respectively. These atmospheric parameters are inherently dependent random variables.

An important part of mission planning is being able to provide, ahead of time, a good assessment of GO/NOGO status for different atmospheric parameters as well as conditional probabilities involving GO and/or NOGO outcomes. Specifically, it is of interest to effectively address certain questions pertaining to the assigned constraints for the different mission phases of the space vehicle (see Smith, Batts, and Willet (1982), also Smith, and Batts (1993)). The questions of interest involve:

1.) The probability that the assigned atmospheric constraints will (or will not) occur during a particular time.

2.) The probability that the assigned atmospheric constraints will (or will not) occur for N consecutive days, at a particular time of the day.

3.) Given that the assigned constraints have occurred (or have not occurred) for n consecutive days, at a particular time of the day, what is the probability that the constraints will continue for N additional days?

4.) The probabilities of runs of GO and NOGO outcomes.

5.) Estimating certain conditional probabilities involving GO and/or NOGO outcomes.

Effectively addressing and giving specific answers to these types of questions are useful in many ways, for instance,

a) determining design criteria for the space vehicle,

b) establishing flight operational rules, and

c) setting up effective cost assessments.

The purpose of this technical report is to present an analytical study of the topics involved, such as the theories of runs and Markov chains, as well as an attempt to give answers to the questions raised above. We construct probabilistic models based on the nature of the problem, as well as certain assumptions relevant to atmospheric conditions. These models lay the ground work to establish a theory that would support a GO - NOGO decision.

II. PRELIMINARIES

The random variables generated by the atmospheric conditions, such as wind speed, thunderstorm, and precipitation, are not independent in general. In fact, a meteorological observation is not usually independent of the preceding conditions. However, the dependence decreases as the length of the time interval between successive events increases. For example, the amount of rain in a month is influenced to a small but definite extent by the amount of rain in the preceding month, but the amount of rain in a year bears practically no relation to the amount of rain in the preceding year. In daily observations the interdependence is found to be even more marked. For example, the probability of a given day being rainy is much greater if it was raining on the preceding day than if it did not. This is due to the fact that rain tends to persist from day to day. Thus, in general, it is the characteristic of meteorological events to stick together; high or low values tend to occur in clusters rather than as isolated incidents.

Brooks and Carruthers (1953) suggested the existence of an underlying Markov chain, however, it does not seem to have been investigated in their work. A Markov chain model for daily rainfall occurrence was used by K. R. Gabriel and J. Neumann (1957, and 1962). This model was shown in their work to give a good fit to various aspects of rainfall occurrence patterns.

The underlying dependence structure in the model is a crucial aspect in the development of our study. Based on the nature of the meteorological observations, and the above mentioned works, it seems reasonable to utilize a dependence structure of the Markovian type.

We should point out here that several authors used the geometric distribution, the negative binomial or related distributions as models for meteorological activities and wet-dry cycles of rain. For example, a distribution of weather cycles by length was investigated using geometric distribution (see K. R. Gabriel, and J. Neumann (1957)). The negative binomial and a modification of it were used as prospective models to represent the variation of thunderstorm activity (see L. W. Falls, W. O. Williford, and M. C. Carter (1970)).

It is interesting to note that these distributions do arise under the Markovian dependence structure which is adopted in our models. Of course, they may also originate under different circumstances.

Next, we give a definition of Markov chains.

1. Markov chains: Consider a sequence of random variables X_0 , X_1 ,..., and suppose that the set of possible values of these random variables is $\{0, 1, \ldots, M\}$. It will be helpful to interpret X_n as being the state of some system at time n, and, in accordance with this interpretation, we say that the system is in state i at time n if $X_n = i$.

The sequence of random variables is said to form a Markov chain if each time the system is in state i there is some fixed probability p_{ij} that it will next be in state j. That is, for all $i_0, \ldots, i_{n-1}$, $i, j \in \{0, \ldots, M\}$

$$P\left\{X_{n+1} = j \mid X_n = i,\ X_{n-1} = i_{n-1}, \ldots,\ X_1 = i_1,\ X_0 = i_0\right\} = p_{ij} . \quad (1)$$

The values p_{ij}, $0 \le i \le M$, $0 \le j \le M$, are called the transition probabilities of the Markov chain and they satisfy

$$p_{ij} \ge 0, \qquad \sum_{j=0}^{M} p_{ij} = 1, \qquad i = 0, 1, \ldots, M. \quad (2)$$

It is convenient to arrange the transition probabilities p_{ij} in a square array $\mathbb{P}$ as follows:

$$\mathbb{P} = \begin{bmatrix} P_{00} & P_{01} & \cdots & P_{0M} \\ P_{10} & P_{11} & \cdots & P_{1M} \\ \vdots & & & \\ P_{M0} & P_{M1} & \cdots & P_{MM} \end{bmatrix}$$

Such an array is called a transition probability matrix. Knowledge of the transition probability matrix and the distribution of X_0 enables us, at least in theory, to compute all probabilities of interest. For instance, the joint probability mass function of $X_0, X_1, \ldots, X_n$ is given by

$$
\begin{aligned}
P\{X_n &= i_n,\ X_{n-1} = i_{n-1}, \ldots, X_1 = i_1,\ X_0 = i_0\} \\
&= P\{X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}\, P\{X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} \\
&= p_{i_{n-1},\, i_n}\, P\{X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\},
\end{aligned}
$$

and continual repetition of this argument yields that the above is equal to

$$
\begin{aligned}
P\{X_n &= i_n,\ X_{n-1} = i_{n-1}, \ldots, X_1 = i_1,\ X_0 = i_0\} \\
&= p_{i_{n-1},\, i_n}\, p_{i_{n-2},\, i_{n-1}} \cdots p_{i_1, i_2}\, p_{i_0,\, i_1}\, P\{X_0 = i_0\}. \qquad (3)
\end{aligned}
$$

Thus, for a Markov chain, p_{ij} represents the probability that a system in state i will enter state j at the next transition. We can also define the two-stage transition probability, $p_{ij}^{(2)}$, that a system, presently in state i, will be in state j after two additional transitions. That is,

$$p_{ij}^{(2)} = P\{X_{m+2} = j \mid X_m = i\}.$$

Then $p_{ij}^{(2)}$ can be computed from the p_{ij} as follows:

$$
\begin{aligned}
p_{ij}^{(2)} &= P\{X_2 = j \mid X_0 = i\} \\
&= \sum_{k=0}^{M} P\{X_2 = j \mid X_1 = k,\ X_0 = i\}\, P\{X_1 = k \mid X_0 = i\} \\
&= \sum_{k=0}^{M} p_{kj}\, p_{ik}\,. \qquad (4)
\end{aligned}
$$

In general, the n-stage transition probabilities, denoted as $p_{ij}^{(n)}$, are defined by

$$p_{ij}^{(n)} = P\{X_{n+m} = j \mid X_m = i\}. \qquad (5)$$

1.2 Remarks: 1.) We note that if p_{ij} is the i-jth entry of the transition matrix $\mathbf{P}$, then $p_{ij}^{(n)}$ is the i-jth entry of $\mathbf{P}^n$.

2.) In the model, a Markov chain with two states will be used. That is, $M = 1$.

1.3 Example: Suppose that if it rains today, then it will rain tomorrow with probability α; and if it does not rain today, then it will rain tomorrow with probability β. If we say that the system is in state 0 when it rains and state 1 when it does not, then the above is a two state Markov chain having transition probability matrix

$$\mathbb{P} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}$$

If it is raining today the probability that it will rain two days from now is

$$p_{00}^{(2)} = p_{00}^2 + p_{01}\, p_{10} = \alpha^2 + (1-\alpha)\beta.$$

The remaining probabilities $p_{01}^{(2)}$, $p_{10}^{(2)}$, and $p_{11}^{(2)}$, which are the entries of $\mathbf{P}^2$, can be given similar interpretations.

$\square$

The next result, known as the Chapman - Kolmogorov equations, shows how the $p_{ij}^{(n)}$ can be computed.

1.4 Lemma (The Chapman - Kolomogorov Equations):

$$p_{ij}^{(n)} = \sum_{k=0}^{M} p_{ik}^{(r)} \quad p_{kj}^{(n-r)} \qquad \text{for all } 0 < r < n .$$

Proof:

$$
\begin{aligned}
p_{ij}^{(n)} &= P\{X_n = j \mid X_0 = i\} \\
&= \sum_k P\{X_n = j,\ X_r = k \mid X_0 = i\} \\
&= \sum_k P\{X_n = j \mid X_r = k,\ X_0 = i\} \quad P\{X_r = k \mid X_0 = i\} \\
&= \sum_k p_{ki}\ p_{ik}\ .
\end{aligned}
$$

□

2. Unconditional probabilities: The conditional probabilities $p_{ij}^{(n)}$ can be used to derive expressions for unconditional probabilities by conditioning on the initial state. For instance,

$$
\begin{aligned}
P\{X_n = j\} &= \sum_{k=0}^{M} P\{X_n = j \mid X_0 = k\}\ \ P\{X_0 = k\} \\
&= \sum_{k=0}^{M} p_{ij}^{(n)} \qquad P\{X_0 = k\}.
\end{aligned} \tag{6}
$$

3. Convergence of Markov Chains: For a large number of Markov chains it turns out that $p_{ij}^{(n)}$ converges, as $n \to \infty$, to a value π_j that depends only on j. That is, for large values of n, the probability of being in state j after n transitions is approximately equal to π_j, no matter what the initial state was. It can be shown that a sufficient condition for a Markov chain to possess this property is that

$$
p_{ij}^{(n)} > 0 \qquad \text{for all } i, j = 0, 1, \ldots, M, \text{ and some } n > 0. \tag{7}
$$

Markov chains that satisfy inequality (6) are said to be ergodic. Since Lemma (2.4) yields

$$
p_{ij}^{(n+1)} \quad = \sum_{k=0}^{M} p_{ik}^{(n)}\ p_{kj}\ ,
$$

it follows, by letting $n \to \infty$, that for ergodic Markov chains

$$\pi_j = \sum_{k=0}^{M} \pi_k \; p_{kj} \, . \tag{8}$$

Furthermore, since $1 = \sum_{j=0}^{M} p_{ij}^{(n)}$, by letting $n \to \infty$, we also obtain,

$$\sum_{j=0}^{M} \pi_j = 1. \tag{9}$$

In fact, it can be shown that the π_j , $0 \leq j \leq M$, are the unique nonnegative solutions of equations (7) and (8). We sum up these facts in the following result, which is stated without proof.

3.1 Theorem: For an ergodic Markov chain the limit $\pi_j = \lim_{n\to\infty} p_{ij}^{(n)}$ exists, and the π_j , $0 \leq j \leq M$ are the unique nonnegative solutions of
$\pi_j = \sum_{k=0}^{M} \pi_k \; p_{kj}$ and $\sum_{j=0}^{M} \pi_j = 1.$

3.2 Example: Consider Example 1.3, in which we assume that if it rains today, then it will rain tomorrow with probability α; and, if it does not rain today, then it will rain tomorrow with probability β. From Theorem 3.1 it follows that the limiting probabilities of rain and of no rain, π_0 and π_1 , are given by

$$\pi_0 = \alpha \, \pi_0 + \beta \pi_1$$

$$\pi_1 = (1 - \alpha) \, \pi_0 + (1 - \beta) \, \pi_1$$

$$\pi_1 + \pi_1 = 1$$

which yields

$$\pi_0 = \frac{\beta}{1 + \beta - \alpha} , \quad \pi_1 = \frac{1 - \alpha}{1 + \beta - \alpha}$$

For instance, if $\alpha = .6$, $\beta = .3$, then the limiting probability of rain on the nth day is $\pi_0 = \frac{3}{7}$.

□

The quantity π_j is also equal to the long run proportion of time that the Markov chain is in state j, $j = 0, \ldots, M$. To intuitively see why this might be so, let p_j denote the long run proportion of time the chain is in state j. Using the strong law of large numbers, it can be shown that for an ergodic chain such long run proportions exist and are constants. Now, since the proportion of time the chain is in state k is p_k and since, when in state k, the chain goes to state j with probability p_{kj}, it follows that the proportion of time the Markov chain is entering state j from state k is equal to $p_k \, p_{kj}$. Summing over all k shows that p_j, the proportion of time the Markov chain is entering state j, satisfies

$$p_j = \sum_{k=0}^{M} p_k \; p_{kj} \; .$$

Since it is also true that

$$\sum_{j=1}^{M} p_j = 1 \, ,$$

it thus follows, since by Theorem 3.1 the π_j, $j = 0, \ldots, M$ are the unique solution of the above, that $p_j = \pi_j$, $j = 0, \ldots, M$.

III. PROBABILISTIC MODELS

The binomial model, which is the sum of independent random variables each of which assumes either a success or a failure outcome, is clearly not adequate for the problem under consideration because it deals with independent trials. So, the need for a model with dependent trials is identified.

We consider in this section a generalization of the binomial distribution which incorporates a built-in dependence structure in its trials. Two versions of a generalized Bernoulli model are presented. Each of these versions possesses two parameters. Therefore, methods of estimating these parameters are discussed including an analytical method based on a modified version of the maximum likelihood estimation technique.

1. A generalization of the Bernoulli model: Consider a sequence of random variables each of which takes either the value 0 or 1 (for GO or NOGO, respectively). We do not assume that these random variables are independent. The basic assumption about this interdependence is that given the present state, the future and the past states are independent. That is,

$$P(X_{i+1}\ X_{i-1} \mid X_i) = P(X_{i+1} \mid X_i)\ P(X_{i-1} \mid X_i). \tag{10}$$

Specifically, given the state (or value) of a random variable at present time i, say X_i, its state (or value) at times $i + 1$, X_{i+1}, depends only on its state (or value) at time i, X_i. This is a Markovian type dependence structure, and such a probabilistic model is referred to as a Markov chain with two states $\{0, 1\}$. A Markov probability model, with two states, usually possesses two parameters. These two parameters could be defined in different ways. For instance, we consider here two versions of this model, in the first we take two parameters as the conditional probabilities

$$\theta_1 = P(X_i = \text{NOGO} \mid X_{i-1} = \text{NOGO}),$$

and

$$\theta_0 = P(X_i = \text{NOGO} \mid X_{i-1} = \text{GO}).$$

In the second version we define the two parameters as

$$p = P(X_i = \text{NOGO}),$$

and

$$\lambda = P(X_i = \text{NOGO} \mid X_{i-1} = \text{NOGO}).$$

Of course, the parameters θ_1, θ_0 in model 1 are related to the parameters p, λ in model 2. Hereafter, we shall use 1 for a NOGO state, and 0 for a GO state.

1.1 Model 1: Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables each of which takes on either the value of 1 or 0 as in the Bernoulli model. A Markovian dependence is incorporated between successive observations. This yields a Markov chain with two states whose parameters are the two conditional probabilities

$$\theta_1 = P(X_i = 1 | X_{i-1} = 1), \quad i = 2, ..., n, \tag{11}$$

and

$$\theta_0 = P(X_i = 1 | X_{i-1} = 0), \quad i = 2, ..., n. \tag{12}$$

From (10) and (11), it follows that

$$P(X_i = 0 | X_{i-1} = 1) = 1 - \theta_1 , \quad i = 2, ..., n, \tag{13}$$

and

$$P(X_i = 0 | X_{i-1} = 0) = 1 - \theta_0 , \quad i = 2, ..., n. \tag{14}$$

The transition probability matrix takes the form

$$\mathbf{P} = \begin{bmatrix} 1-\theta_0 & \theta_0 \\ 1-\theta_1 & \theta_1 \end{bmatrix} \tag{15}$$

Clearly, this model reduces to the case of independent Bernoulli trials if $\theta_0 = \theta_1$. Explicit formulas for the n-stage transition probabilities $p_{ij}^{(n)}$ can be obtained using the transition probability matrix $\mathbf{P}$ (see Feller, 1957). The final result may be written in matrix form, the n-stage transition matrix

$$\mathbf{P}^n = \frac{1}{1 + \theta_0 - \theta_1} \begin{bmatrix} 1-\theta_1 & \theta_0 \\ 1-\theta_1 & \theta_0 \end{bmatrix} + \frac{(\theta_1 - \theta_0)^n}{1 + \theta_0 - \theta_1} \begin{bmatrix} \theta_0 & -\theta_0 \\ -(1-\theta_1) & 1-\theta_1 \end{bmatrix}. \tag{16}$$

In the transition matrix $\mathbb{P}^n$ each element $p_{ij}^{(n)}$, $i, j = 0, 1,$ represents the probability that a system, presently in state i, will be in state j after n additional transitions.

1.2 Remarks: (i) We observe here that, since $|\theta_1 - \theta_0| < 1$, the second matrix in the definition of $\mathbb{P}^n$ in (7) tends to zero as $n \to \infty$. That is,

$$\lim_{n\to\infty} \mathbb{P}^n = \frac{1}{1+\theta_0-\theta_1}\begin{bmatrix} 1-\theta_1 & \theta_0 \\ 1-\theta_1 & \theta_0 \end{bmatrix}$$

$$= \frac{1}{1+\theta_0-\theta_1}\,\mathbb{P}. \tag{17}$$

This result is consistent with the result obtained had we used Theorem 3.1.

(ii) This model requires estimating the two conditional probabilities θ_1 and θ_0. One way to estimate θ_1 and θ_0 is by using the appropriate relative frequencies (see K. R. Gabriel and J. Neumann 1962). An analytical method, based on the maximum likelihood estimation technique, will be described later.

(iii) A Markov chain model of this form was found to be, at least, a close approximation for the daily rainfall occurrence at Tel Aviv. It was shown also that it fits the observed data (see K. R. Gabriel and J. Neumann 1962).

1.3 Model 2: This model differs from model 1 in that one of its parameters has a different meaning. Consider a sequence $X_1, X_2, \ldots, X_n$ of random variables each of which takes on either the value 1 or 0. This model possesses the usual frequency parameter

$$p = P(X_i = 1) = 1 - P(X_i = 0), \qquad i = 1, 2, \ldots, n, \tag{18}$$

and, in addition, another parameter θ which measures the dependence or the degree of persistence in the chain.

$$\begin{aligned} \theta &= P(X_i = 1 \mid X_{i-1} = 1) \\ &= 1 - P(X_i = 0 \mid X_{i-1} = 1), \qquad i = 2, 3, \ldots, n. \end{aligned} \tag{19}$$

The relationship between model 2 and model 1 can be obtained, using (18) and (19), by computing the other conditional probabilities as follows:

Let $x = P(X_i = 1 \mid X_{i-1} = 0)$, then

$$\begin{aligned} p &= P(X_i = 1,\ X_{i-1} = 1) + P(X_i = 1,\ X_{i-1} = 0) \\ &= P(X_i = 1 \mid X_{i-1} = 1)\ P(X_{i-1} = 1) + P(X_i = 1 \mid X_{i-1} = 0)\ P(X_{i-1} = 0) \\ &= \theta\ p + x(1 - p), \end{aligned}$$

from which

$$x = P(X_i = 1 \mid X_{i-1} = 0) = \frac{(1-\theta)p}{1-p} \tag{20}$$

and

$$P(X_i = 0 \mid X_{i-1} = 0) = \frac{1 - 2p + \theta p}{1 - p}\,. \tag{21}$$

Thus $\{X_1, \ldots, X_n\}$ is a stationary two-state Markov chain with transition matrix

$$\mathbf{P} = \begin{bmatrix} \dfrac{1 - 2p + \theta p}{1 - p} & \dfrac{(1-\theta)p}{1-p} \\ 1-\theta & \theta \end{bmatrix}. \tag{22}$$

In this case the n-stage transition matrix takes the form

$$\mathbf{P}^n = \begin{bmatrix} 1-p & p \\ 1-p & p \end{bmatrix} + \left(\frac{\theta - p}{1 - p}\right)^n \begin{bmatrix} p & -p \\ -(1-p) & 1-p \end{bmatrix}. \tag{23}$$

Using Theorem 3.1 it can be shown that $\pi_0 = 1 - p$ and $\pi_1 = p$, which is what we expect in this case.

1.4 Remark: A way to estimate the parameters p and θ is by using the appropriate relative frequency, as in the case with model 1. However, there is an analytical way to estimate these parameters. We describe this method next.

2. Estimating the parameters: As pointed out earlier, the parameters of either model can be estimated by using the appropriate relative frequency, see example, K.R. Gabriel and J. Neumann 1962. Here we describe an analytical method based on the idea of modifying the maximum likelihood (see Billingsley (1961), also see Devore (1976)).

2.1 The modified maximum likelihood method:

Let the transition matrix be $\mathbf{P} = \begin{bmatrix} 1-\theta_0 & \theta_0 \\ 1-\theta_1 & \theta_1 \end{bmatrix}$.

The full likelihood (or the joint distribution of $X_1, \ldots, X_n$), using the Markov assumptions in terms of θ_0 and θ_1, can be written as

$$P(X_1 = x_1,\ X_2 = x_2, \cdots,\ X_n = x_n)$$

$$= P(X_1 = x_1)\cdot P(X_2 = x_2 | X_1 = x_1)\cdot P(X_3 = x_3 | X_2 = x_2) \cdots P(X_n = x_n | X_{n-1} = x_{n-1})$$

$$= p^{x_1}\ (1-p)^{1-x_1} \prod_{i=2}^{n} \theta_1^{x_{i-1}\ x_i}\ (1-\theta_1)^{(x_{i-1})(1-x_i)}$$
$$\times\ \theta_0^{(1-x_{i-1})x_i}\ (1-\theta_0)^{(1-x_{i-1})(1-x_i)},$$

where $p = P(X_i = 1) = \dfrac{\theta_0}{1+\theta_0-\theta_1}$.

The term $p^{x_1}(1-p)^{1-x_1}$ represents the contribution due to the first state visited by the process. The modified maximum likelihood method consists of neglecting the term $p^{x_1}(1-p)^{1-x_1}$. This idea was used by P. Billingsley in his development of the asymptotic theory of maximum likelihood estimators, (see P. Billingsley (1961)). Therefore, this method is particularly useful when n is large.

Now, the natural log of the likelihood, denoted by L, for the realization $x_1, x_2, \ldots, x_n$ can be expressed as

$$L = x_1 \ln p + (1 - x_1) \ln(1-p) + L', \tag{24}$$

where

$$L' = n_{00} \ln(1-\theta_0) + n_{01} \ln\theta_0 + n_{10} \ln(1-\theta_1) + n_{11} \ln \theta_1 . \tag{25}$$

The n_{ij} , i, j = 0, 1, are the usual transition counts given by the number of indices m for which $x_m = i$ and $x_{m+1} = j$, $m = 1, \ldots, n-1$ so that $n_{00} + n_{01} + n_{10} + n_{11} = n - 1$. For example, n_{11} is the number of indices m such that $x_m = 1$ and $x_{m+1} = 1$, for $m = 1, \ldots, n-1$. Therefore, the required estimators of θ_0 and θ_1 are obtained by maximizing L′ given in (25).

$$\frac{\partial L'}{\partial \theta_0} = \frac{n_{00}}{1-\theta_0}(-1) + \frac{n_{01}}{\theta_0} = 0 ,$$

which yields

$$\hat{\theta}_0 = \frac{n_{01}}{n_{00} + n_{01}} \tag{26}$$

$$\frac{\partial L'}{\partial \theta_1} = \frac{n_{10}}{1-\theta_1}(-1) + \frac{n_{11}}{\theta_1} = 0 ,$$

which yields

$$\hat{\theta}_1 = \frac{n_{11}}{n_{10} + n_{11}} . \tag{27}$$

2.2 Remarks: (i) We note that $\frac{\partial^2 L'}{\partial \theta_0^2} < 0$ and $\frac{\partial^2 L'}{\partial \theta_1^2} < 0$, so, the values $\hat{\theta}_0$ and $\hat{\theta}_1$, given by (26) and (27) do indeed maximize L'.

(ii) In the above derivation, both θ_0 and θ_1 are free to vary, so that when the value of θ_0 is specified, the modified maximum likelihood of θ_1 is no longer given by (27).

(iii) For model 2, we can get the estimators of p and θ by comparing the entries in (15) and (22), to obtain

$$\hat{\theta} = \hat{\theta}_1 = \frac{n_{11}}{n_{10} + n_{11}} , \tag{28}$$

and $\frac{(1-\theta_1)p}{1-p} = \theta_0$, which yields $p = \frac{\theta_0}{1 + \theta_0 - \theta_1}$.

Therefore,

$$\hat{p} = \frac{\theta_0}{1 + \hat{\theta}_0 - \hat{\theta}_1} , \tag{29}$$

where $\hat{\theta}_0$ and $\hat{\theta}_1$ are given in (26) and (27), and it simplifies after some algebra to

$$\hat{p} = \frac{n_{10} \, (n_{10} + n_{11})}{n_{00} \, n_{10} + 2 \, n_{01} \, n_{10} + n_{01} \, n_{11}} . \tag{30}$$

IV. DISTRIBUTIONS OF GO-NOGO STATES

The two models presented in section 3 are technically equivalent. Therefore, for convenience, we only use model 1 to discuss some distributions which involve the number of GO and/or NOGO states.

1. Distribution of the number of states of the same kind:

1.1 Distribution of the number of GO states: Let X be the number of successive GO states until a NOGO state occurs. Then,

$$\underbrace{00\ldots0}_{n}1$$

$$P(X = n) = (1 - \theta_0)^{n-1} \quad \theta_0 \ , \quad n = 1, 2, \ldots \tag{31}$$

This is a geometric or Pascal probability distribution (a special case of the negative binomial).

1.2 Distribution of the number of NOGO states: Similarly, if Y is the number of successive NOGO states until a GO state occurs. Then,

$$\underbrace{11\ldots1}_{m}0$$

$$P(Y = m) = \theta_1^{m-1} \ (1 - \ \theta_1) \ , \ m = 1, 2, \ldots \tag{32}$$

2. Distribution of recurrence time: A succession of NOGO states of length k, $k \geq 0$, means a sequence of k-NOGO states preceded and followed by GO states. Therefore, a NOGO succession of length k is equivalent to a recurrence time of k+1 for GO states.

2.1 Distribution of recurrence time of NOGO states: The following table illustrates how the probabilities of different recurrence of NOGO states are computed.

k	Recurrence time of NOGO	Representation	Probability
0	1	1 1	θ_1
1	2	1 0 1	$(1-\theta_1)\theta_0$
2	3	1 0 0 1	$(1-\theta_1)(1-\theta_0)\theta_0$
3	4	1 0 0 0 1	$(1-\theta_1)(1-\theta_0)^2\theta_0$
$\vdots$			
k	k+1	$1\overbrace{0\,0 \ldots 0}^{k}1$	$(1-\theta_1)(1-\theta_0)^{k-1}\theta_0$

2.2 Lemma: The mean and variance of recurrence time T_1 of NOGO states are given by

$$\mu_1 = \frac{1-(\theta_1-\theta_0)}{\theta_0}, \tag{33}$$

and

$$\sigma_1^2 = \frac{1-\theta_1}{\theta_0^2}(1+\theta_1-\theta_0) \tag{34}$$

Proof: $$\mu_1 = 1\theta_1 + 2(1-\theta_1)\theta_0 + 3(1-\theta_1)(1-\theta_0)\theta_0 + \cdots$$

$$= \theta_1 + \frac{(1-\theta_1)}{1-\theta_0}\theta_0\,[\,2(1-\theta_0) + 3(1-\theta_0)^2 + \cdots\,]$$

$$= \theta_1 + \frac{(1-\theta_1)}{1-\theta_0}\theta_0(-1)\,\frac{d}{d\theta}\,[(1-\theta_0)^2 + (1-\theta_0)^3 + \cdots\,]$$

$$= \theta_1 + \frac{(1-\theta_1)}{1-\theta_0}\theta_0(-1)\,\frac{d}{d\theta}\,[\,\frac{(1-\theta_0)^2}{\theta_0}\,]$$

$$= \theta_1 + \frac{(1-\theta_1)}{1-\theta_0}\theta_0(-1) \quad [\frac{-1}{\theta_0^2} + 1]$$

$$= \theta_1 + \frac{(1-\theta_1)(1+\theta_0)}{\theta_0}$$

$$= \frac{1-(\theta_1-\theta_0)}{\theta_0} \; .$$

Now,

$$\sigma_1^2 = E(T_1^2) - \mu_1^2 \; , \quad \text{where}$$

$$E(T_1^2) = 1\,\theta_1 + 2^2(1-\theta_1)\theta_0 + 3^2(1-\theta_1)(1-\theta_0)\theta_0 + \cdots$$

$$= \theta_1 + \frac{(1-\theta_1)}{1-\theta_0}\theta_0 \; [\, 2^2(1-\theta_0) + 3^2(1-\theta_0)^2 + \cdots \,]$$

$$= \theta_1 + \frac{(1-\theta_1)}{1-\theta_0}\theta_0(-1)\frac{d}{d\theta_0}[2(1-\theta_0)^2 + 3(1-\theta_0)^3 + \cdots \,]$$

$$= \theta_1 + \frac{(1-\theta_1)}{1-\theta_0}\theta_0(-1)\frac{d}{d\theta_0}[(1-\theta_0)\,(2(1-\theta_0) + 3(1-\theta_0)^2 + \cdots)]$$

$$= \theta_1 + \frac{(1-\theta_1)}{1-\theta_0}\theta_0(-1)\frac{d}{d\theta_0}[(1-\theta_0)\,(\frac{1}{\theta_0^2} - 1)]$$

$$= \theta_1 - \frac{(1-\theta_1)}{1-\theta_0}\theta_0[(1-\theta_0)\,(\frac{-2}{\theta_0^3}) + \frac{1-\theta_0^2}{\theta_0^2}(-1)]$$

$$= \theta_1 + \frac{2(1-\theta_1)}{\theta_0^2} + \theta_0\frac{(1-\theta_1)(1+\theta_0)}{\theta_0^2}$$

$$= \frac{2(1-\theta_1) + \theta_0 + \theta_0^2 - \theta_0\theta_1}{\theta_0^2}$$

Thus

$$\sigma_1^2 = \frac{2(1-\theta_1)+\theta_0 + \theta_0^2 - \theta_0\theta_1}{\theta_0^2} - \frac{1-2(\theta_1-\theta_0)+(\theta_1-\theta_0)^2}{\theta_0^2}$$

$$= \frac{1-\theta_0+\theta_0\theta_1-\theta_0^2}{\theta_0^2}$$

$$= \frac{(1-\theta_1)}{\theta_0^2}(1+\theta_1-\theta_0).$$

□

2.3 Distribution of recurrence time of GO states. We illustrate this also by a table

k	Recurrence time of GO	Representation	Probability
0	1	100	$1-\theta_1$
1	2	010	$\theta_0(1-\theta_1)$
2	3	0110	$\theta_0\theta_1(1-\theta_1)$
3	4	01110	$\theta_0\theta_1^2(1-\theta_1)$
⋮			
k	k+1	$0\overset{k}{11\ldots1}0$	$\theta_0\theta_1^2(1-\theta_1)$

2.4 Lemma: The mean and variance of recurrence time T_0 of GO states are given by

$$\mu_0 = \frac{1-(\theta_1-\theta_0)}{1-\theta_0}, \tag{35}$$

and

$$\sigma_0^2 = \frac{\theta_0}{(1-\theta_1)^2}(1+\theta_1-\theta_0). \tag{36}$$

Proof. Is similar to the proof of Lemma 2.2, hence we omit it.

□

3. Asymptotic distribution of the number of NOGO/GO states: Since NOGO and GO states are represented by 1 and 0, respectively, it follows that the total number of NOGO states in a sequence of n trials $\{X_1, X_2, \ldots, X_n\}$ is $S = \sum_{k=1}^{n} X_k$. It is shown in Feller, 1957, Chapter XIII, that S can be approximated by a normal distribution, provided that n is large enough, this fact is given in the next result without proof.

3.1 Theorem (Normal approximation of recurrence time of NOGO states): If the recurrence time of NOGO states has mean μ_1 and variance σ_1^2, then the number of NOGO states in n trials $S = \sum_{k=1}^{n} X_k$ is asymptotically normally distributed with approximate mean $\frac{n}{\mu_1}$ and approximate variance $\frac{n\sigma_1^2}{\mu_1^3}$, where μ_1 and σ_1^2 are given by (33) and (34).

A similar result for the recurrence time of GO states is given next.

3.2 Theorem (Normal approximation of recurrence time of GO states): If the recurrence time of GO states has mean μ_0 and variance σ_0^2, then the number of GO states in n trials is asymptotically normally distributed with approximate mean and variance $\frac{n}{\mu_0}$ and $\frac{n\sigma_0^2}{\mu_0^3}$, respectively, where μ_0 and σ_0^2 are given by (35) and (36).

3.3 Remark: The asymptotic results, Theorems 3.1 and 3.2, neither tell how rapid the distributions approach normality nor reflect the exact distributions for small n. The exact distribution will be discussed in section V, analysis of runs.

4. The distribution of numbers of successive GO followed by successive NOGO states:

Let X be a random variable representing the number of successive GO states and let Y be a random variable representing the number of successive NOGO states. Then X and Y are independent random variables. Define $Z = X + Y$. Then, Z represents a number of successive GO states followed by a number of successive NOGO states, and

$$P\{Z = n\} = \sum_{k=1}^{n-1} P(X = k)\, P(Y = n - k)$$

$$= \sum_{k=1}^{n-1} \theta_1^{k-1} (1 - \theta_1)(1 - \theta_0)^{n-k-1} \theta_0$$

$$= \theta_0 (1 - \theta_1) \sum_{k=1}^{n-1} \theta_1^{k-1} (1 - \theta_0)^{n-k-1}$$

$$= \theta_0 (1 - \theta_1) \frac{(1 - \theta_0)^{n-1} - \theta_1^{n-1}}{(1 - \theta_0) - \theta_1} . \tag{37}$$

We note that the distribution (7) is symmetric in θ_0 and $(1 - \theta_1)$, this is expected because $Z = X + Y = Y + X$.

4.1 Remarks: (i) The distribution (37) can be considered as a generalization of the negative binomial in the following sense. Let $1 - \theta_0 = \theta_1$, then

$$\lim_{(1-\theta_0)\to\theta_1} P\{Z = n\} = \theta_0 (1 - \theta_1)(n - 1)\theta_1^{n-2}$$

$$= (n - 1)(1 - \theta_1)^2 \theta_1^{n-2} ,$$

which is the negative binomial probability of having the second NOGO to occur at the nth trial.

(ii) The work of L. W. Falls, W. O. Williford, M. C. Carter, 1970, reached the conclusion that the negative binomial and a modification of the negative binomial distribution are adequate statistical models to represent thunderstorm events and thunderstorm hits, respectively, at Cape Kennedy, Florida.

V. ANALYSIS OF RUNS

We consider here sequences of NOGO and GO states. Here we follow the theory of run as in S. S. Wilks, 1963, however our work provides some generality as the probability of a GO outcome is not necessarily equal to that of a NOGO outcome. Suppose that the total number of GO states is n_0 and that of the NOGO states is n_1, with $n_0 + n_1 = n$. The class of all these sequences is actually the set of all $\binom{n}{n_0}$ permutations of n_0 GO states, and n_1 NOGO states. Each sequence consists of runs of NOGO and GO states. The length of a run is the number of states in it.

Let r_{0k} denote the number of runs of GO (or 0) states of length k, and let r_{1j} denote the number of runs of NOGO (or 1) states of length j. For example; the sequence 00011001001101 is such that $n_0 = 8$, $n_1 = 6$, $r_{01} = 1$, $r_{02} = 2$, $r_{03} = 1$, $r_{11} = 2$, $r_{12} = 2$, and all other r's being zero.

From the definition of these quantities we see that $\sum_k k\, r_{0k} = n_0$ and $\sum_j j\, r_{1j} = n_1$. Let $r_0 = \sum_k r_{0k}$, and $r_1 = \sum_j r_{1j}$ be the total number of runs of GO and NOGO states, respectively. Given the set of values of the r_{0k}, the number of ways of arranging the r_0 runs of GO states is the multinomial coefficient

$$\binom{r_0}{r_{01}, r_{02}, \ldots, r_{0n_0}} = \frac{r_0!}{r_{01}!\, r_{02}! \ldots r_{0n_0}!} . \tag{38}$$

Likewise, the number of ways of arranging the r_1 runs of NOGO states is

$$\binom{r_1}{r_{11}, r_{12}, \ldots, r_{1n_1}} = \frac{r_1!}{r_{11}!\, r_{12}! \ldots r_{1n_1}!} . \tag{39}$$

1. Remark: We observe here that r_0 and r_1 can differ from each other by at most 1; because if they differ by more than 1, this means that at least two runs of one kind of states would have to be adjacent, contradicting the definition of a run. If $r_0 = r_1$, then there are two ways of arranging the runs of GO and NOGO states, one sequence begins with a run of GO state(s) and the other begins with a run of NOGO state(s).

2. Definition: Let $\delta(r_0, r_1)$ be the number of ways of arranging r_0 indistinguishable objects of one kind and r_1 indistinguishable objects of a second kind such that no two objects of the same kind appear together, then $\delta(r_0, r_1)$ takes two possible values

$$\delta(r_0, r_1) = \begin{cases} 1, & \text{if } |r_0 - r_1| = 1 \\ 2, & \text{if } |r_0 - r_1| = 0 \end{cases} . \tag{40}$$

Therefore, the total number of ways of having r_{0k} runs of GO states of lengths $k = 1, 2, \ldots, n_0$ and of having r_{1j} runs of NOGO states of lengths $j = 1, 2, \ldots, n_1$ is

$$\frac{r_0!\, r_1!\ \delta(r_0, r_1)}{r_{01}!\, r_{02}! \cdots r_{0n_0}!\ r_{11}!\, r_{12}! \cdots r_{1n_1}!} . \tag{41}$$

3. The probability of a fixed sequence of GO-NOGO states:

Consider a fixed sequence with r_{0k} runs of GO states of lengths $k = 1, 2, \ldots, n_0$ and r_{1j} runs of NOGO states of lengths $j = 1, 2, \ldots, n_1$. In order to obtain the required probability, call it $P(E)$, we start by conditioning on the state of the first trial. That is, letting H denote the event that the first trial results in a NOGO, we then have

$$P(E) = p\,P(E|H) + (1-p)\,P(E|H^c) \tag{42}$$

where $p = P(NOGO) = \dfrac{\theta_0}{1 + \theta_0 - \theta_1}$, see equation (20) with $x = \theta_1$. Given that the initial trial is a NOGO, then

$$P(E|H) = \theta^{n_1 - r_1} \; (1-\theta_0)^{n_0 - r_0} \; (1-\theta_1)^{r_1 + \delta - 2} \; \theta_0^{r_0 - \delta + 1}\,.$$

where $\theta_1^{n_1 - r_1}$ accounts for all the consecutive NOGO (or 1) states, $(1 - \theta_0)^{n_0 - r_0}$ accounts for all the consecutive GO (or 0) states, the term $(1 - \theta_1)^{r_1 + \delta - 2}$ accounts for the number of changes from a NOGO to a GO (or from 1 to 0) state. We note that if $r_0 = r_1$, then $\delta = 2$ and there are r_1 changes from NOGO to GO state. On the other hand, if $r_1 > r_0$, then $\delta = 1$ and there are $r_1 - 1$ changes from NOGO to GO state. The case $r_1 < r_0$ can not occur, by definition of runs since we begin with a run of NOGO states. Finally, the term $\theta_0^{r_0 - \delta + 1}$ accounts for the number of changes for a GO to a NOGO (or from 0 to 1) state. We observe here that if $r_0 = r_1$, then $\delta = 2$ and there are $r_0 - 1$ changes from a GO to a NOGO state, while there are r_0 changes from a GO to a NOGO state if $r_1 > r_0$.

Thus

$$\begin{aligned} P(E|H) &= \theta_1^{n_1 - r_1} \; (1 - \theta_0)^{n_0 - r_0} \; (1 - \theta_1)^{r_1 + \delta - 2} \; \theta_0^{r_0 - \delta + 1} \\ &= \theta_1^{n_1 - r_1} \; (1 - \theta_0)^{n_0 + 1} \left(\frac{1 - \theta_1}{\theta_1}\right)^{r_1 - 2} \left(\frac{\theta_0}{1 - \theta_1}\right)^{r_0 + 1} \left(\frac{1 - \theta_1}{\theta_0}\right)^{\delta}. \end{aligned} \tag{43}$$

Similarly, given that the initial state is a GO state, then

$$
\begin{aligned}
P(E|H^c) &= (1-\theta_0)^{n_0 - r_0} \quad \theta_1^{n_1 - r_1} \quad \theta_0^{r_0 + \delta - 2} \quad (1-\theta_0)^{r_1 - \delta + 1} \\
&= (1-\theta_0)^{n_0 - 2} \quad \theta_1^{n_1 + 1} \left(\frac{\theta_0}{1-\theta_0}\right)^{r_0 - 2} \left(\frac{1-\theta_1}{\theta_1}\right)^{r_1 + 1} \left(\frac{\theta_0}{1-\theta_1}\right)^{\delta} .
\end{aligned}
\tag{44}
$$

Therefore, substituting (43) and (44) in (42), the probability P(E) of a given sequence with r_{0k} runs of GO states of lengths $k = 1, 2, \ldots, n_0$ and r_{1j} runs of NOGO states of lengths $j = 1, 2, \ldots, n_1$ is

$$
\begin{aligned}
P(E) &= p\theta_1^{n_1 - 2} \; (1-\theta_0)^{n_0 + 1} \left(\frac{1-\theta_1}{\theta_1}\right)^{r_1 - 2} \left(\frac{\theta_0}{1-\theta_0}\right)^{r_0 + 1} \left(\frac{1-\theta_1}{\theta_0}\right)^{\delta} \\
&+ (1-p)\,(1-\theta_0)^{n_0 - 2} \; \theta_1^{n_1 + 1} \left(\frac{\theta_0}{1-\theta_0}\right)^{r_0 - 2} \left(\frac{1-\theta_1}{\theta_1}\right)^{r_1 + 1} \left(\frac{\theta_0}{1-\theta_1}\right)^{\delta} ,
\end{aligned}
\tag{45}
$$

where $p = \dfrac{\theta_0}{1 + \theta_0 - \theta_1}$.

Since the event E in the above analyses may occur a total of $\dfrac{r_0 !\, r_1 !\;\; \delta(r_0 , r_1)}{r_{01}! \cdots r_{0n_0}!\;\; r_{11}! \cdots r_{1n_1}!}$ ways, it follows that the probability of a sequence of NOGO and GO states, as described above is

$$
\frac{r_0 !\, r_1 !\;\; \delta(r_0 , r_1)}{r_{01}! \cdots r_{0n_0}!\;\; r_{11}! \cdots r_{1n_1}!} \quad P(E) := P(\{r_{ij}\}). \tag{46}
$$

4. Probability of runs of GO states of length k, $k = 1, 2, \ldots, n_0$:

The probability formula obtained in (46) can be thought of as a joint p.d.f. of runs of NOGO and GO states. If we are interested only in the p.d.f. of the runs of GO states, that is, r_{0k}, $k = 1, 2, \ldots, n_0$, we have to take the marginal distribution of $(r_{01}, r_{02}, \ldots, r_{0n_0})$ in (46). That is, we must sum the probability $P(\{r_{ij}\})$ with respect to $r_{11}, r_{12}, \ldots, r_{1n_1}$. This means that we must sum the formula (46) for all values r_{11}, $r_{12}, \ldots, r_{1n_1}$ such that $\sum_j jr_{1j} = n_1$ and $\sum_j r_{1j} = r_1$. In order to do this we make use of the following identity in x which holds for values of x near zero:

$$
\begin{aligned}
(x + x^2 + x^3 + \cdots)^{r_1} &\equiv x^{r_1} (1 - x)^{-r_1} \\
&\equiv x^{r_1} \sum_{i=0}^{\infty} \frac{(r_1 + i - 1)!}{(r_1 - 1)!\, i!}\, x^i . \qquad (47)
\end{aligned}
$$

Now, the coefficient of x^{n_1} in the first expression of the sum of $\dfrac{r_1!}{r_{11}! \cdots r_{1n_1}!}$ with respect to the $r_{11}, r_{12}, \cdots, r_{1n_1}$ subject to the restrictions $\sum_j jr_{1j} = n_1$ and $\sum_j r_{1j} = r_1$. But the coefficient of x^{n_1} from the first expression in (47) must equal that of x^{n_1} in the last expression in (47) which is seen to be

$$
\frac{(n_1 - 1)!}{(r_1 - 1)!\,(n_1 - r_1)!} = \binom{n_1 - 1}{r_1 - 1}. \qquad (48)
$$

Hence, the p.d.f. of runs of GO states r_{0k}, $k = 1, 2, \ldots, n_0$, and a total number of NOGO states equal to r_1 is

$$
P(\{r_{0k}\}, r_1) = \frac{r_0!}{r_{01}! \cdots r_{0n_0}!} \binom{n_1 - 1}{r_1 - 1} \delta(r_0, r_1)\, P(E), \qquad (49)
$$

where $P(E)$ is as given in (45).

Finally, to obtain the p.d.f of the r_{0k} , $k = 1,2, \ldots, n_0$, we must sum (49) with respect to r_1 . Using the definition of $\delta(r_0, r_1)$ we see that

$$\sum_{r_1=1}^{n_1} \binom{n_1-1}{r_1-1} \delta(r_0, r_1)\, p(E)$$

$$= 2\binom{n_1-1}{r_0-1} P(E)|_{r_1=r_0} + \binom{n_1-1}{r_0} P(E)|_{r_1=r_0+1} + \binom{n_1-1}{r_0-2} P(E)|_{r_1=r_0-1} \tag{50}$$

$$:= \alpha\ .$$

Therefore,

$$P\big(\{r_{0k}\}\big) = \frac{r_0!\, \alpha}{r_{01}! \ldots r_{0n_0}!}, \tag{51}$$

where $k = 1,2,\ldots, n_0$, and α is defined by (50).

5. Probability of runs of NOGO states of length $k = 1,2,\ldots, n_1$:

By similar analyses we see that

$$P\big(\{r_{1k}\}\big) = \frac{r_1!\, \beta}{r_{11}! \cdots r_{1n_1}!}, \tag{52}$$

where $k = 1,2,\ldots, n_1$, and

$$\beta := \sum_{r_0=1}^{n_0} \binom{n_0-1}{r_0-1} \delta(r_0, r_1)\, P(E) \tag{53}$$

6. The joint distribution of the total number of runs of GO states and the total number of runs of NOGO states:

The p.d.f of r_0 and r_1 can be obtained by summing (49) with respect to r_{0k} subject to the conditions that $\sum_k k\, r_{0k} = n_0$ and $\sum_k r_{0k} = r_0$.
The technique here is similar to that used in summing (46) to obtain (49) and it yields

$$P(r_0, r_1) = \binom{n_0 - 1}{r_0 - 1}\binom{n_1 - 1}{r_1 - 1}\delta(r_0, r_1)\, P(E). \tag{54}$$

7. The distribution of the total number of runs:

The p.d.f. of r_0, the total number of runs of GO states, $P(r_0)$ can be obtained by summing (54) with respect to r_1. The p.d.f. of r_1, the total number of runs of NOGO states $P(r_1)$ is obtained in a similar way.

VI. CONDITIONAL PROBABILITIES

The models presented in section III can be used to answer certain conditional probability questions in a straight forward way. Suppose we know that GO states occurred for the past n days, at a particular time of the day, what is the probability that the GO state will continue for N additional days?

Based on the Markovian property utilized in the models the question may be stated as follows: given that a GO state has occurred, at a particular time of the day, what is the probability that the GO state will continue for N additional days? Symbolically, the situation can be represented as $\underset{N}{000\cdots0}$

Therefore,

$$P(GO = N|GO) = (1 - \theta_0)^N . \tag{55}$$

Similarly, we have $\underset{N}{100\cdots0}$

$$P(GO{=}N|NOGO) = (1 - \theta_1)(1 - \theta_0)^{N-1} , \tag{56}$$

$$P(NOGO{=}N|GO) = \theta_0 \theta_1^{N-1} , \qquad \underset{N}{011\cdots1} \tag{57}$$

and

$$P(NOGO{=}N|NOGO) = \theta_1^N . \qquad \underset{N}{11\cdots1} \tag{58}$$

VII. EXAMPLE

To illustrate the methods discussed in this report we give an example. We use the 33 years of data for thunderstorm and non-thunderstorm days for summer (months of June, July, and August), see Smith and Batts, Tables 13a and 13b. For comparison purposes we include the same tables here as tables 1 and 2.

a) Parameter estimation:

To estimate the parameter θ_1, θ_0 we use formulas (26) and (27):

$$\hat{\theta}_1 = \frac{n_{11}}{n_{10} + n_{11}} \qquad \hat{\theta}_0 = \frac{n_{01}}{n_{00} + n_{01}}$$

where n_{11} is the frequency of having two consecutive NOGO days (i.e., two back to back thunderstorm days), n_{00} is the frequency of having two consecutive GO days (i.e., two back to back no thunderstorm days), n_{01} is the frequency of having a GO day followed by a NOGO day, and n_{10} is the frequency of having a NOGO day followed by a GO day.

From table 1: $n_{11} = 859$, and $n_{01} = 1363$.

From table 2: $n_{00} = 1328$, and $n_{10} = 1848$.

Therefore,

$$\hat{\theta}_1 = \frac{n_{11}}{n_{10} + n_{11}} = \frac{859}{1848 + 859} = 0.3173 ,$$

and

$$\hat{\theta}_0 = \frac{n_{01}}{n_{00} + n_{01}} = \frac{1363}{1328 + 1363} = 0.5065.$$

The significance of the estimates $\widehat{\theta}_1$, $\widehat{\theta}_0$ is that we actually summarized the 33 years of data in just two numbers !!

We observe here that in estimating the conditional probabilities θ_1 and θ_0 we used both tables, i.e. the data for thunderstorm days as well as the data for non-thunderstorm days! This is because the thunderstorm days and the non-thunderstorm days are "elusively correlated." An important feature of the methods in this report is that the correlation between GO and NOGO days is taken into consideration. This is the basic difference between these methods and the method suggested by O. E. Smith and G. W. Batts.

To estimate the parameter p , we use formula (29) or (30)

$$\widehat{p} = \frac{\widehat{\theta}_0}{1 + \widehat{\theta}_0 - \widehat{\theta}_1} = \frac{0.5065}{1 + 0.5065 - 0.3173} = \frac{0.5065}{1.1892} = 0.4259 \doteq 42.6\%,$$

where, $\widehat{p}$ represents the probability of NOGO conditions.

(b) Time conditional probabilities:

Using the estimates $\widehat{\theta}_1$, $\widehat{\theta}_0$, we can compute several probabilities of interest mentioned in this report. For example,

(1.) The probability of n successive GO states until a NOGO state occurs (formula 31) is

$$0.5065 \, (1 - 0.5065)^{n-1} \quad , \qquad n = 1, 2, \ldots.$$

(2.) The probability of m successive NOGO states until a GO state occurs (formula (42)) is

$$(1 - 0.3173)\,(\,0.3173\,)^{m-1}$$
$$= 0.6827\,(\,0.3173\,)^{m-1}\quad,\quad m = 1, 2, \ldots.$$

Formulas (55), (56), (57), and (58) yield:

(3.) Given that a GO day has occurred, the probability that the GO state will continue for N additional days is

$$P(GO = N \mid GO) = (\,1 - 0.5065\,)^{N}\quad,$$

Similarly,

(4.) $$P(GO = N \mid NOGO) = 0.6827\,(\,1 - 0.5065\,)^{N-1}\,,$$

(5.) $$P(NOGO = N \mid GO) = 0.5065\,(\,0.3173\,)^{N-1}\,,$$

and

(6.) $$P(NOGO = N \mid NOGO) = (\,0.3173\,)^{N}.$$

where $N = 1, 2, 3 \cdots$.

REFERENCES

1. C. E. P. Brooks, N. Carruthers: Handbook of Statistical Methods in Meteorology, London, Her Majesty's Stationery Office (1953).

2. P. Billingsley: Statistical Inference for Markov Processes, Univ. of Chicago Press (1961).

3. J. L. Devore: A note on the estimation of parameters in a Bernoulli model with dependence. The Ann. Statist., **4** #5 (1976), 990-992.

4. L. W. Falls, W. O. Williford, and M. C. Carter: Probability distributions for thunderstorm activity at Cape Kennedy, FL, NASA Tech. Memorandum, TM X-53867 (1970).

5. W. Feller: An Introduction to Probability Theory and its Applications, 2nd ed., 1, New York (Wiley) (1957).

6. K. R. Gabriel, and J. Newman: A Markov chain model for daily rainfall occurrence at Tel Aviv. Quart. J. Roy. meteorological Soc., **88** (1962), 90-95.

7. K. R. Gabriel: The distribution of the number of successes in a sequence of dependent trials. Biometrika, **46** (1959), 454-460.

8. K. R. Gabriel, and J. Newman: On a distribution of weather cycles by length. Quart. J. Roy. meteorological Soc., **83** (1957), 375-380.

9. J. Klotz: Statistical Inference in Bernoulli trials with dependence. The Ann. Statist., **1** (1973), 373-379.

10. O. E. Smith, G. W. Batts, and J. A. Willett: Atmospheric constraint statistics for the space shuttle mission planning, NASA Tech. Paper #2069 (1982).

11. O. E. Smith, and G. W. Batts: Atmospheric statistics for the aerospace vehicle operations, AIAA 31st Aerospace Sciences Meeting, AIAA 93-0756 (January 1993).

12. S. S. Wilks: Mathematical Statistics, 2nd ed., New York (Wiley) (1963).

TABLE 1. Distribution of Runs and Time Conditional Probabilities for Thunderstorm Days for Summer, KSC

THUNDERSTORM DAYS AT K S C (1957-1989)
MONTHS OF JUNE, JULY, AUGUST

LENGTH	FREQ	SUM 1	SUM 2	PROB.	J=1	J=2	J=3	J=4	J=5	J=6	J=7	J=8	J=9	J=10
1	206	504	1363	0.4245	1.0000									
2	118	298	859	0.2675	0.6302	1.0000								
3	56	180	561	0.1747	0.4116	0.6531	1.0000							
4	41	124	381	0.1187	0.2795	0.4435	0.6791	1.0000						
5	28	83	257	0.0800	0.1886	0.2932	0.4581	0.6745	1.0000					
6	14	55	174	0.0542	0.1277	0.2026	0.3102	0.4567	0.6770	1.0000				
7	18	41	119	0.0371	0.0873	0.1385	0.2121	0.3123	0.4630	0.6839	1.0000			
8	6	23	78	0.0243	0.0572	0.0908	0.1390	0.2047	0.3035	0.4483	0.6555	1.0000		
9	7	17	55	0.0171	0.0404	0.0640	0.0980	0.1444	0.2140	0.3161	0.4622	0.7051	1.0000	
10	3	10	38	0.0118	0.0279	0.0442	0.0677	0.0997	0.1479	0.2184	0.3193	0.4872	0.6909	1.0000
11	0	7	28	0.0087	0.0205	0.0326	0.0499	0.0735	0.1089	0.1609	0.2353	0.3590	0.5091	0.7368
12	2	7	21	0.0065	0.0154	0.0244	0.0374	0.0551	0.0817	0.1207	0.1765	0.2692	0.3818	0.5526
13	1	5	14	0.0044	0.0103	0.0163	0.0250	0.0367	0.0545	0.0805	0.1176	0.1795	0.2545	0.3684
14	1	4	9	0.0028	0.0066	0.0105	0.0160	0.0236	0.0350	0.0517	0.0756	0.1154	0.1636	0.2368
15	1	3	5	0.0016	0.0037	0.0058	0.0089	0.0131	0.0195	0.0287	0.0420	0.0641	0.0909	0.1316
16	2	2	2	0.0006	0.0015	0.0023	0.0036	0.0052	0.0078	0.0115	0.0168	0.0256	0.0364	0.0526

TABLE 2. Distribution of Runs and Time Conditional Probabilities for NON-Thunderstorm Days for Summer, KSC

NO THUNDERSTORM DAYS AT K S C (1957-1989)
MONTHS OF JUNE, JULY, AUGUST

LENGTH	FREQ	SUM 1	SUM 2	PROB.	J=1	J=2	J=3	J=4	J=5	J=6	J=7	J=8	J=9	J=10
1	180	520	1848	0.5755	1.0000									
2	85	340	1328	0.4136	0.7186	1.0000								
3	67	255	988	0.3077	0.5346	0.7440	1.0000							
4	48	188	733	0.2283	0.3966	0.5520	0.7419	1.0000						
5	41	140	545	0.1697	0.2949	0.4104	0.5516	0.7435	1.0000					
6	26	99	405	0.1261	0.2192	0.3050	0.4099	0.5525	0.7431	1.0000				
7	15	73	306	0.0953	0.1656	0.2304	0.3097	0.4175	0.5615	0.7556	1.0000			
8	14	58	233	0.0726	0.1261	0.1755	0.2358	0.3179	0.4275	0.5753	0.7614	1.0000		
9	14	44	175	0.0545	0.0947	0.1318	0.1771	0.2387	0.3211	0.4321	0.5719	0.7511	1.0000	
10	5	30	131	0.0408	0.0709	0.0986	0.1326	0.1787	0.2404	0.3235	0.4281	0.5622	0.7486	1.0000
11	5	25	101	0.0315	0.0547	0.0761	0.1022	0.1378	0.1853	0.2494	0.3301	0.4335	0.5771	0.7710
12	4	20	76	0.0237	0.0411	0.0572	0.0769	0.1037	0.1394	0.1877	0.2484	0.3262	0.4343	0.5802
13	5	16	56	0.0174	0.0303	0.0422	0.0567	0.0764	0.1028	0.1383	0.1830	0.2403	0.3200	0.4275
14	2	11	40	0.0125	0.0216	0.0301	0.0405	0.0546	0.0734	0.0988	0.1307	0.1717	0.2286	0.3053
15	2	9	29	0.0090	0.0157	0.0218	0.0294	0.0396	0.0532	0.0716	0.0948	0.1245	0.1657	0.2214
16	3	7	20	0.0062	0.0108	0.0151	0.0202	0.0273	0.0367	0.0494	0.0654	0.0858	0.1143	0.1527
17	2	4	13	0.0040	0.0070	0.0098	0.0132	0.0177	0.0239	0.0321	0.0425	0.0558	0.0743	0.0992
18	0	2	9	0.0028	0.0049	0.0068	0.0091	0.0123	0.0165	0.0222	0.0294	0.0386	0.0514	0.0687
19	1	2	7	0.0022	0.0038	0.0053	0.0071	0.0095	0.0128	0.0173	0.0229	0.0300	0.0400	0.0534
20	0	1	5	0.0016	0.0027	0.0038	0.0051	0.0068	0.0092	0.0123	0.0163	0.0215	0.0286	0.0382
21	0	1	4	0.0012	0.0022	0.0030	0.0040	0.0055	0.0073	0.0099	0.0131	0.0172	0.0229	0.0305
22	0	1	3	0.0009	0.0016	0.0023	0.0030	0.0041	0.0055	0.0074	0.0098	0.0129	0.0171	0.0229
23	0	1	2	0.0006	0.0011	0.0015	0.0020	0.0027	0.0037	0.0049	0.0065	0.0086	0.0114	0.0153
24	1	1	1	0.0003	0.0005	0.0008	0.0010	0.0014	0.0018	0.0025	0.0033	0.0043	0.0057	0.0076

APPROVAL

PARAMETRIC ANALYSIS OF ATMOSPHERIC PROCESSES

By

M. Elshamy

This report has been reviewed for technical accuracy and contains no information concerning national security or nuclear energy activities or programs. The report, in its entirety, is unclassified.

for Samuel H. Morgan

Gregory S. Wilson
Director, Space Science Laboratory

☆ U.S. GOVERNMENT PRINTING OFFICE 1993–733050/80111